

Public Response to Air Pollution on Indonesian Social Media

**Aulia Hasanah Nur Rahman^{*1}, Muhammad Khatami Sahl², Salsabila Shafiyah³,
Darrell Nathaniel Pieter Gees Manua⁴, Nadia Qudsia⁵, Fajri Maulana⁶, Resa Septiani
Pontoh⁷**

^{1, 2, 3, 4, 5, 6, 7} Department of Statistics, Universitas Padjadjaran, Indonesia.

(E-mail: aulia21011@mail.unpad.ac.id, muhammad21280@mail.unpad.ac.id,
salsabila21010@mail.unpad.ac.id, darrell21001@mailunpad.ac.id, nadia21013@mail.unpad.ac.id,
fajri21002@mail.unpad.ac.id, resa.septiani@mail.unpad.ac.id)

ABSTRACT

Air pollution, a critical global environmental concern impacting individual and community health, is investigated in this study using the BERT (Bidirectional Encoder Representations from Transformers) model to classify sentiments as positive, negative, or neutral in Indonesia. Analysis involves data sources derived from replies and posts made on X, offering insights into public perceptions and air quality concerns. The research explores prevailing sentiments associated with air pollution, classifying them into positive, negative, or neutral. To assess the BERT-based sentiment analysis model, the F1 Score, a widely recognized metric that balances precision and recall, is utilized. Initially, an imbalanced model yielded an accuracy ranging from 66-67%, while a balanced model achieved an accuracy of 72%. Despite limitations in hardware, the initial findings show promising potential in uncovering valuable insights into the public's attitudes towards air pollution in Indonesia. This study provides a comprehensive analysis of public perceptions of air pollution, forming a basis for future research and environmental interventions in Indonesia.

Keywords: BERT, Indonesia, Natural Language Processing, Pollution, Sentiment Analysis

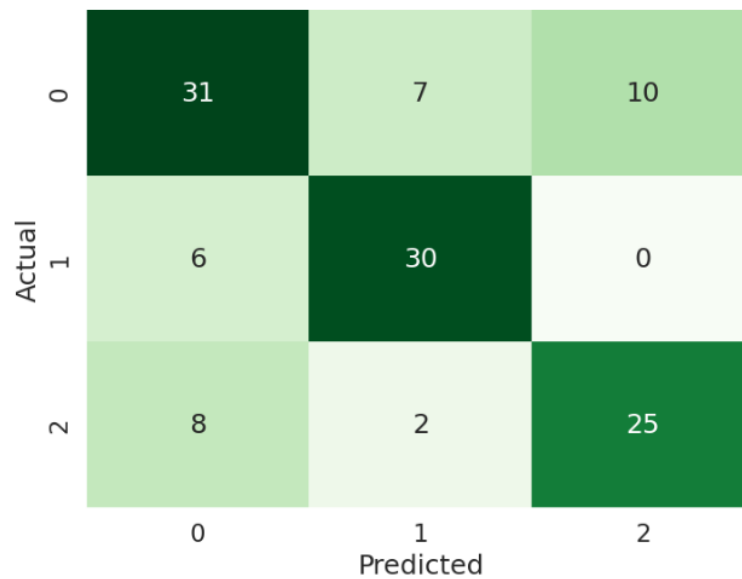


Figure 1. Confusion Matrix for BERT with Balanced Training

Table 1. Sentiment Classification Results for BERT with Balanced Training

	Precision	Recall	F1-Score
Neutral	0.69	0.65	0.67
Positive	0.77	0.83	0.80
Negative	0.71	0.71	0.71
Accuracy			0.72
Weighted avg	0.72	0.72	0.72

References:

- [1] Rosenberg, E., Tarazona, C., Mallor, F., et al.: ‘Sentiment analysis on Twitter data towards climate action’Results in Engineering, 2023, 19, p. 101287.
- [2] Catelli, R., Pelosi, S., Comito, C., Pizzuti, C., Esposito, M.: ‘Lexicon-based sentiment analysis to detect opinions and attitude towards covid-19 vaccines on Twitter in Italy’Computers in Biology and Medicine, 2023, 158, p. 106876.
- [3] Sun, Y., Ji, M., Jin, F., Wang, H.: ‘Public responses to air pollution in Shandong province using the online complaint data’ISPRS International Journal of Geo-Information, 2021, 10, (3), p. 126.
- [4] Jiao, Y., Li, C., Lin, Y.: ‘Can urban environmental problems be accurately identified? A complaint text mining method’Applied Sciences, 2021, 11, (9), p. 4087.
- [5] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
- [6] Habbat, N., Nouri, H., Anoun, H., Hassouni, L.: ‘Sentiment analysis of imbalanced datasets using Bert and ensemble stacking for Deep Learning’Engineering Applications of Artificial Intelligence, 2023, 126, p. 106999.